Where in London should Starbucks open a new coffee shop?

Alan Sun

July 10, 2021

1 Introduction

1.1 Background

Starbucks is a chain coffee store, and a staple of London's high streets, with shops all over the City. As with every chain store, Starbucks is always looking for opportunities and locations to open new shops.

1.2 Problem

This report will provide analysis and advice to the Starbucks stakeholders as to where in London they should open up their newest coffee shop. (Please don't sue me for defamation - all companies mentioned in this notebook are fictitious, and likeness to any existing companies is purely coincidental).

Greater London is already very naturally divided into 33 principal divisions - The 32 London boroughs, as well as the City of London. So we will use these divisions to decide which would be the most profitable borough in which to open Starbucks' newest branch. This decision will be based on Foursquare location data for each borough, and the venue types that feature most frequently in each borough.

To give a crude example, it could be the case that people who visit coffee shops also eat at Italian restaurants. Then you would expect that it would be the case that a new coffee shop opened in an area with a high density of Italian restaurants would be more profitable than one opened in an area with very few Italian restaurants.

2 Data acquisition and cleaning

2.1 Data sources

The data concerning the locations of each of the London Boroughs is taken from here. This was combined with Foursquare location date for each borough in order to produce our table of features.

2.2 Data cleaning

Data scraped from the above Wikipedia page was cleaned using BeautifulSoup, and put into a pandas dataframe. The data for the City of London had to be added manually, as it is not technically a borough.

2.3 Feature selection

After data cleaning, Foursquare location data was used to retrieve the venue types of the venues within 3 kilometres of the borough location. The venue types that accounted for more than 1% of the total venues were selected as the feature variables. The number of coffee shops in each borough was the target variable.

3 Exploratory Data Analysis

I used Folium to create a map of London, in order to visualise the Geography and location of each of the boroughs.

After the location data was obtained through Foursquare, Onehot encoding was used to group the venues in each borough by venue type, and the total number of each venue type in each borough were then used as the feature variables in our model.

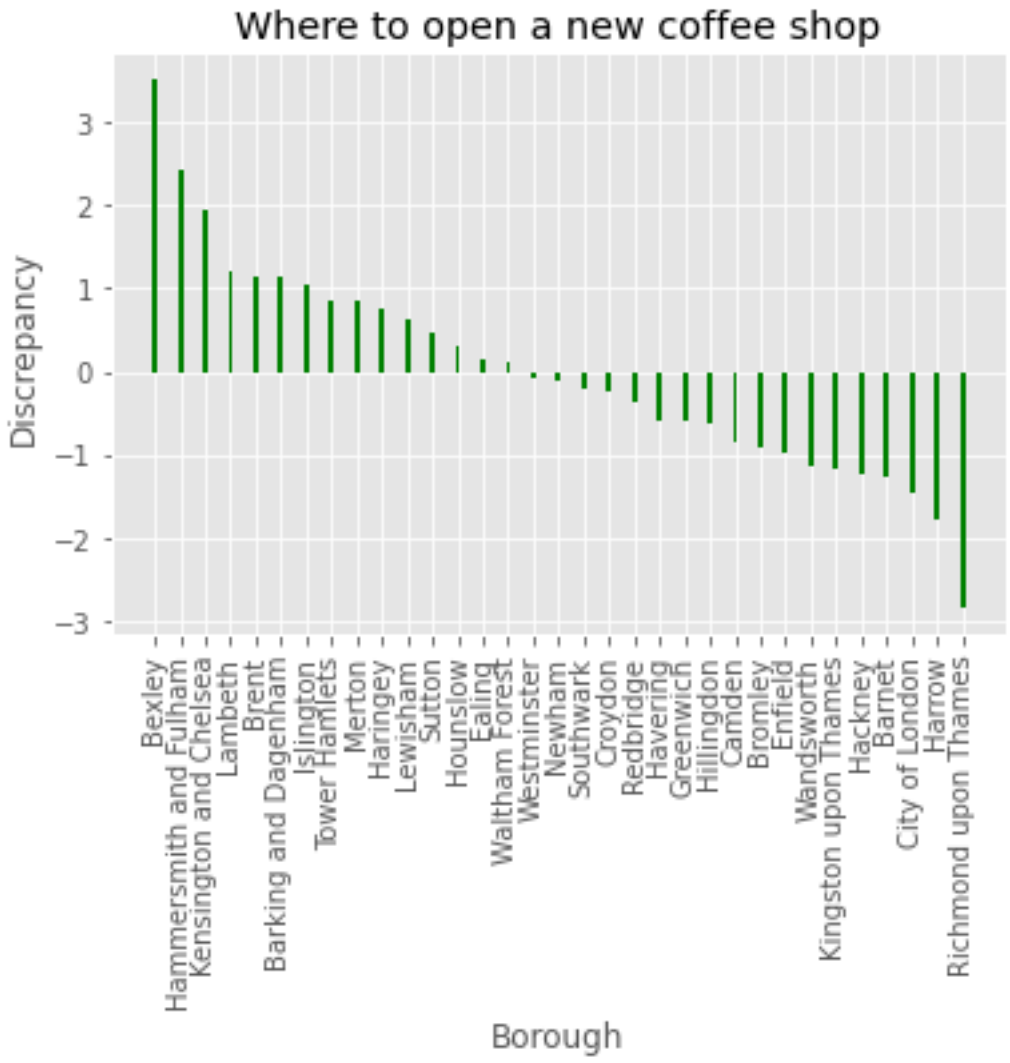
4 Predictive Modelling

4.1 Linear Regression Model

We employed a linear model to see if it the number of other venue types in the borough could predict the number of coffee shops in the borough.

The heights of the bars in the bar chart below represent the the difference between the actual number of coffee shops in each borough, and the predicted number of coffee shops in each borough, i.e. the discrepancy. For example, a tall bar with positive height, represents a borough in which there are far fewer coffee shops than we would otherwise expect, given the frequency of the other venue types. So in theory, the boroughs on the x axis of the below chart are

arranged in order, with the boroughs furthest to the left being the ones with the best potential in which to open a new coffee shop.



5 Discussion

Of course, there are many limitations to this approach. The first of which is the linear assumption - why should there be such a relationship between different venue types located around the same place? If there is such a relationship

between the venue types in each borough, it is almost certainly a higher order relationship, that can't be quantified by such a simple linear model.

Secondly, even in the case that our linear assumption is valid, must it be that a lower predicted number of coffee shops than actual coffee shops implies that an extra coffee shop in the area would be profitable? This is predicated on the idea that the current supply (the number of coffee shops in the borough) is dwarfed by the current demand (the number of coffee shops we expect to see in the borough based on the types of the other venues present), and so incrementally increasing the supply of coffee shops in this borough will result in the greatest expected profit.

Lastly, there is the argument that a chain as large as Starbucks need not worry about such issues. They have the resources to open many, many new chains all over the City, and as long as each one is profitable (or at least breaks even), then there is not necessarily any need for them to focus on the potential profitability of each branch, taking the "quantity over quality" approach that works for so many large firms.

6 Conclusions

In this study, I looked at the venue types in each borough, and used them to try and predict the best locations in which a new coffee shop would most likely thrive.

Our linear model suggests that Bexley, Hammersmith and Fulham, and Kensington and Chelsea would be the 3 boroughs in which a new Starbucks store would be most likely to make the largest profit. As discussed above though, these suggestions should perhaps be taken with a pinch of salt, as this rudimentary study is a huge over-simplification of the true dynamics of a huge urban sprawl such as London.